DOCUMENT RESUME

ED 157 871 95 SP 012 743

AUTHOR Marliave, Richard S.
TITLE A Review of the Findings of Phase II. Technical Note I-1. Beginning Teacher Evaluation Study.
INSTITUTION Far West Lab. for Educational Research and Development, San Francisco, Calif.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE May 76
CONTRACT 400-65-0001
NOTE 38p.; For related documents, see SP 012 741-751

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS *Academic Achievement; Academic Aptitude; Curriculum Planning; *Educational Research; Learning Processes; *Student Behavior; *Student Characteristics; Teacher Background; *Teacher Behavior; *Teacher Characteristics

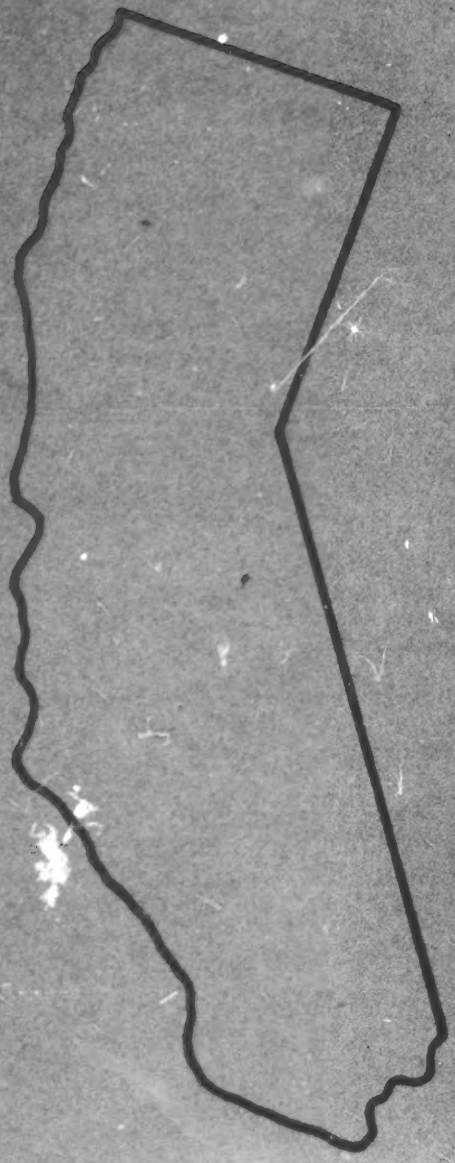
ABSTRACT

This document examines the implications of the Phase II report specifically for observation variables that might be used in future Beginning Teacher Evaluation Study (BTES) research, focusing on those variables and patterns of variables that showed the strongest and most consistent relationships with student learning in the Phase II report. The variables are discussed within four basic categories: (1) teacher behaviors; (2) student behaviors; (3) teacher characteristics; and (4) student characteristics. This document is written with the assumption that the reader is familiar with the design and methodology of the Phase II study. The Phase II report examines the data collection, data analysis, and conclusions of one of the most comprehensive and complex studies of teaching conducted to date. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED157871

Aug 21 10 59 AM '78



BTES

BEGINNING TEACHER EVALUATION STUDY

Technical Note Series

TECHNICAL NOTE I-1

A Review of the Findings of Phase II

Richard S. Marliave

May, 1976

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

SP018 743

FAR WEST LABORATORY
FOR EDUCATIONAL RESEARCH AND DEVELOPMENT
1855 FOLSOM STREET ~ SAN FRANCISCO, CALIFORNIA 94103 ~ (415) 565-3000

This work was completed pursuant to contract #400-65-0001 between the Department of Health, Education and Welfare, National Institute of Education and the California Commission for Teacher Preparation and Licensing.

The Far West Laboratory for Educational Research and Development, 1855 Folsom Street, San Francisco, California 94103, is a nonprofit organization supported in part by the United States Office of Education and the National Institute of Education, Department of Health, Education and Welfare. The opinions expressed in this publication do not necessarily reflect the position or policy of these agencies, and no official endorsement by them should be inferred.

The Phase II report[1] of the Beginning Teacher Evaluation Study (BTES) examines the data collection, data analysis, and conclusions of one of the most comprehensive and complex studies of teaching conducted to date. It is not the purpose of this technical note to give a complete summary or a critical review of that report. Rather, the purpose, here, is to examine the implications of the Phase II report specifically for observation variables that might be used in future BTES research. Therefore, this technical note will focus on those variables and patterns of variables that showed the strongest and most consistent relationships with student learning in the Phase II report. These should be the most promising candidates for future research. It should be noted that this technical note is written with the assumption that the reader is familiar with the design and methodology of the Phase II study.

The Phase II variables will be discussed within four basic categories: 1) teacher behaviors, 2) student behaviors, 3) teacher characteristics, and 4) student characteristics. These categories are used here primarily for organizational purposes. These are broad and loosely conceived clusters that the reader might profitably choose to restructure for his/her own purposes. Teacher behavior refers here to instructional process variables that are under the control of the teacher, such as planning activities (of the teacher), grouping patterns (determined by the teacher), materials

[1] McDonald, F. J., Elias, P., Stone, M., Wheeler, P., Lambert, N., Calfee, R., Sandoval, J., Ekstrom, R., & Lockheed, M. Final report on Phase II of the Beginning Teacher Evaluation Study. Prepared for the California Commission for Teacher Preparation and Licensing, Sacramento, California. Princeton, New Jersey: Educational Testing Service, 1975.

(selected by the teacher), and instructional communications (of the teacher). These variables are assessed by observation and teacher self-report. Student behavior involves the learning events of students that occur in the classroom, presumably as a function of the instructional process. These variables are primarily observed variables. There will also be some discussion of the relationships between teacher and student behaviors. Teacher characteristics are variables outside of the instructional process itself, that may be used to explain the origin of patterns of teacher behavior. These variables, including teacher expectations, knowledge, aptitude, etc., are generally not observable. Student characteristics are also not part of the instructional process or learning events, yet are potentially related to student behavior in the classroom. These variables, including student background, aptitude, etc., are also not readily observable.

Related issues will also be discussed. Consideration will be given to potentially useful reanalyses that might be conducted in the Phase II data. Suggested reanalyses range from a restructuring of the data set based on alternative variable definitions to reorganization of analysis procedures as a means of examining alternative relationships between variables. The purpose of reanalysis would be to enhance our understanding of the Phase II data set. There will also be some discussion of the methodological implications of the Phase II results as they bear upon the data sets of future research. That is, what has Phase II taught us about conducting research on teaching? This discussion will focus on the processes of conceptualizing and measuring variables for future research.

Teacher Behaviors

The most striking finding of Phase II is that no teaching skill was significantly related to student learning in each of the four "quadrants" (grade 2 reading, grade 2 mathematics, grade 5 reading, and grade 5 mathematics). This suggests two important conclusions. First, research involving the kinds of teaching skills investigated in Phase II should be designed to identify effects that are specific to grade level and/or subject matter. Second, alternative conceptualizations of teaching skills and behaviors should be considered with the aim of determining whether some more generalizable outcomes can be identified. Regarding this second conclusion, the Phase II report recommends the investigation of patterns of teaching behavior rather than specific skills. This is a major conclusion and will receive further discussion later in this technical note.

The findings for particular teaching skills will be discussed, here, in terms of two levels of analysis. First, results will be considered for the 22 BTES performance variables as analyzed with stepwise regression procedures, reported in Volume I of the Phase II report. Each of these variables is a combination of one or more specific variables from the Work Diary, RAMOS, and APPLE instruments, yielding the 22 more general categories. Hence, the second level of analysis for findings discussed here is that of the specific teaching skill variables as measured by particular instruments. Volume III of the Phase II report includes the results for specific skill variables of RAMOS and APPLE, as well as results for the evaluation of videotape recordings. Specific variable results for the Work Diary are reported in Volume V.

The teaching skills discussed below are organized into four general areas: 1) teacher planning, 2) settings, (whole class, group, and individual

contexts that are teacher-led [illegible] dependent), 3) instructional materials, and 4) teacher interaction with students (subdivided into management and instruction). It should be noted that the bulk of the Phase II findings for teacher behaviors are in the area of teacher interaction, both management and instruction.

Planning. Teacher behaviors in the area of planning and preparation are analyzed with data from the Work Diary instrument. One of the 22 performance variables, Work Diary-1, is the time spent by the teacher in preparing for and teaching reading and mathematics. Roughly half of the time included in this variable is preparation time. The results of stepwise regression analyses relating Work Diary-1 with student learning in each of the four quadrants were consistently negative, except for the grade five mathematics quadrant for which there was generally no effect of Work Diary-1. These generally negative relationships can be attributed to preparation time, rather than instructional time, because other results to be discussed below show a positive relationship between the teacher's instructional time and student learning. The more specific analyses of Work Diary data tend to support this conclusion, in that greater amounts of teacher preparation time are associated with lower student learning in reading, although there was generally no effect for mathematics.

The general relationship discussed above is for the quantity of teacher time devoted to planning and preparation. It seems likely that negative learning outcomes for preparation are related only to the amount of time devoted to preparation, rather than the quality of the teacher's preparation for instruction. In fact, there was some indication that the quality of preparation would be positively related to student learning. Work Diary-3, a performance variable rating the quality of the teaching

methodology as reported in the Work Diaries (primarily for skills introduced and reviewed), showed a generally positive relationship with student learning for grade 2 mathematics outcomes. However, Work Diary-3 showed no general relationship to outcomes in the other three quadrants. Nevertheless, it is reasonable to conclude that whereas a negative relationship was found between teacher preparation time and student learning, preparation itself may still be a valuable activity if considered in terms of its quality. This would suggest that future research on teacher preparation and planning activities should investigate the qualitative, rather than the quantitative (time), characteristics of these activities.

Settings. Teacher behavior in the area of settings established by the teacher (whole class, group, and individual contexts that are teacher-led or independent) are analyzed with data from the Work Diary, RAMOS, and APPLE instruments. Generally, the results seem to be inconsistent and somewhat difficult to interpret when different setting variables are combined for analysis. However, the more specific analyses seem to indicate that teacher-led instruction results in greater student learning than independent student work. A more tentative conclusion is that the more individualized settings result in more student learning. However, it is difficult to separate the teacher-led/independent dimension from the group size (individualization) dimension, because a greater proportion of teacher-led instruction is possible with whole class instruction while more independent student work is necessary with smaller groups and more individualized settings.

The inconsistency of the setting results is most evident with the analysis of Work Diary data. This may be due to the fact that these data are based on teacher self-report (rather than observation), where it may be more difficult to obtain reliable abstractions of particular dimensions

from confounded setting variable patterns. For example, Work Diary-4, a performance variable that primarily reflects the proportion of students taught by the teacher, was associated with greater student learning for grade 5 mathematics outcomes, but lower learning outcomes for both reading and mathematics at grade 2. Analyses of more specific data from the Work Diary instrument showed teacher-led instruction to be associated with higher reading outcomes, especially at grade 2. This was true both for the amount of teaching time and for the proportion of the students taught exclusively by the teacher. A related finding was that lower outcomes occur when adults other than the teacher (such as aides) provided reading instruction. However, this could be a function of the characteristics of students assigned to other adults (for example, those students might often be the slower learners)

The Work Diary results for grouping patterns are also somewhat inconsistent. Whole class instruction was associated with higher reading outcomes at grade 5, but mixed outcomes at grade 2 (higher decoding and lower reading application). Small group instruction was associated with higher reading outcomes at grade 2, lower reading outcomes at grade 5, and lower mathematics outcomes at grade 2. Individual student settings were associated with lower reading outcomes at grade 2 and higher mathematics outcomes at grade 2.

Two RAMOS variables are relevant to setting. RAMOS-1 is a performance variable (combining several more specific variables) that primarily reflects the proportion of instruction led directly the the teacher. Stepwise regression analyses showed RAMOS-1 to be associated with higher mathematics outcomes at grade 2. Effects were not found in the other three quadrants. RAMOS-4 is a performance variable that primarily reflects the amount of time students spend doing seatwork. This was associated with greater learning

of grade 5 reading and lower learning outcomes for mathematics (grades 2 and 5). Seatwork may be less effective for mathematics than for reading, but these results are difficult to interpret because RAMOS-4 does not take into account the level of teacher interaction with students during seatwork. Teacher interaction with students may be crucial to student productivity during seatwork.

Three APPLE performance variables are relevant to setting. APPLE-5, consisting of independent student work, did not result in any general effects when analyzed using stepwise regression. APPLE-6, teacher-led group activities, was associated with higher mathematics outcomes at grade 5 and lower mathematics outcomes at grade 2. APPLE-7, teacher-led whole class instruction, was also associated with higher mathematics outcomes at grade 5 and lower mathematics outcomes at grade 2. Obviously, these results do not suggest many generalizations. However, these three performance variables do not separate the teacher-led/independent dimension from the group size dimension.

Analyses of more specific APPLE variables distinguish between these two dimensions and provide consistent results. Teacher-led instruction was associated with higher reading outcomes at grades 2 and 5 and higher mathematics outcomes at grade 5. Independent whole class or group settings were associated with lower mathematics outcomes at grades 2 and 5, while the independent group setting was associated with lower reading outcomes. Hence, the specific analyses of APPLE data showed consistently higher outcomes for teacher-led instruction as opposed to independent student work. Disregarding the teacher-led/ independent dimension, individual settings were associated with higher reading and mathematics outcomes at grade 2. Hence, the specific APPLE analyses tended to favor individual settings, as opposed to group or whole class settings.

Summarizing the Phase II results for setting, teacher-led instruction seems to be generally associated with higher student learning outcomes. The results for group size are less clear, but there is some indication that more individualized settings are more productive, although this is probably complicated by a tendency for individualized settings to involve more independent student work. It is safe to conclude that future research should attempt to distinguish between the extent to which the teacher is directly involved in instruction and the extent to which the instructional setting is individualized (as opposed to whole class).

Materials. Phase II results in the area of materials examine the effects of the extent to which a variety of materials was used during instruction. The two performance variables used to assess variety of materials are Work Diary-5 and RAMOS-3. There was a consistent negative relationship between variety of materials and student learning in all quadrants except grade 2 reading, where there was a positive relationship. This inconsistency does not lend itself to simple interpretation. However, the generally negative results for variety of materials may be due to the detraction of sheer variety from the structure and focus of an instructional program. That is, the teacher may find it more cumbersome to coordinate a greater variety of materials and make appropriate allocations of different materials to each student. In addition, the students may find it difficult to adjust to the different organizational and substantive characteristics of diverse materials.

The positive results of variety of materials for grade 2 reading would be consistent with the above explanation if grade 2 reading is seen as a quadrant where general practice is especially appropriate and where variety of materials (used in reading practice) would enhance learning transfer to tests of reading skills. That is, structure and focus may not be critical

at grade 2 reading if a program of general practice using diverse materials is seen as an effective instructional procedure. An alternative explanation of the same relationship is that a variety of materials enhances the structure and focus of instruction that is based on general reading practice.

It is at least possible to conclude that variety of materials is usually, but not consistently, associated with lower student achievement. Future research should investigate the role of variety of materials as a variable that mediates the structure and focus of instruction.

Interaction for Management. Several Phase II variables are related to teacher interaction with students for the purpose of classroom management. Three of the performance variables are relevant here. APPLE-8 reflects the proportion of teacher organizing and supervising that occurred during instruction. This was associated with lower reading outcomes at grade 5 and no effects for any of the other three quadrants. APPLE-13 measures the use of social control behaviors that are task orienting. This variable was associated with higher mathematics outcomes at both grades. APPLE-14 assesses the teacher's nonresponsiveness to (ignoring of) individual student behavior. Effects were not found for APPLE-14. These results are difficult to interpret. They may suggest that instruction is more effective when the teacher actively attempts to maintain a task orientation. However, management without a task orientation, responding to an immediate need to control student behavior, may be an indication of generally poor control of behavior and a low level of attention to instruction.

Results for the more specific variables from the APPLE instrument, as well as some results for the evaluation of videotape recordings, show more

consistently lower outcomes for classroom management behaviors.[2] Academic organizing and instruction giving were associated with lower reading outcomes. Disciplining and redirecting were associated with lower mathematics outcomes. The evaluation of videotape recordings showed the teacher behavior, giving directions, to be associated with poorer student learning at grade 5. These results indicate that management behaviors are associated with lower levels of student learning. However, this may be due to the fact that teachers with disorderly classes are likely to spend more time disciplining students and trying to establish order. That is, the teachers who manage less effectively may have to spend more time trying to manage (and less time providing substantive instruction).

Future research should explore management behaviors not only in terms of their frequency, but also in terms of their source. That is, management behaviors that are due to the teacher's general task orientation may be beneficial to instruction. However, management behaviors that are a response to immediate student behavior problems may be symptomatic of poor management and less effective instruction.

Interaction for Instruction. Phase II data are related to several aspects of teacher interaction with students for the purpose of instruction (rather than management). Perhaps the most simple (though by no means

[2]It should be noted that differences between the results for the performance variables (APPLE-8, 13, and 14) and results for more specific variables may be artifacts of properties of the complex statistical procedures employed. For example, multicollinearity among variables entered in the regression analysis of performance variables may have complicated the results to the extent that interpretations, such as those suggested here, are inappropriate.

consistently lower outcomes for classroom management behaviors.[2] Academic organizing and instruction giving were associated with lower reading outcomes. Disciplining and redirecting were associated with lower mathematics outcomes. The evaluation of videotape recordings showed the teacher behavior giving directions, to be associated with poorer student learning at grade 5. These results indicate that management behaviors are associated with lower levels of student learning. However, this may be due to the fact that teachers with disorderly classes are likely to spend more time disciplining students and trying to establish order. That is, the teachers who manage less effectively may have to spend more time trying to manage (and less time providing substantive instruction).

Future research should explore management behaviors not only in terms of their frequency, but also in terms of their source. That is, management behaviors that are due to the teacher's general task orientation may be beneficial to instruction. However, management behaviors that are a response to immediate student behavior problems may be symptomatic of poor management and less effective instruction.

Interaction for Instruction. Phase II data are related to several aspects of teacher interaction with students for the purpose of instruction (rather than management). Perhaps the most simple (though by no means

[2]It should be noted that differences between the results for the performance variables (APPLE-8, 13, and 14) and results for more specific variables may be artifacts of properties of the complex statistical procedures employed. For example, multicollinearity among variables entered in the regression analysis of performance variables may have complicated the results to the extent that interpretations, such as those suggested here, are inappropriate.

trivial) implications of the results is that more instruction, in terms of time, is associated with more learning. RAMOS-1 represents the amount of time devoted to direct instruction, including the facilitation and monitoring of instruction. This performance variable was positively related to grade 2 mathematics outcomes, although no effects were found in the other quadrants. Analyses of specific APPLE variables (not performance variables) showed direct instruction by the teacher to be positively associated with reading outcomes. The only negative results for instructional time were found for the performance variable, Work Diary-1, as discussed previously in this technical note. However, these negative results are probably due to the fact that about half of the time included in Work Diary-1 was preparation time, rather than classroom instructional time. Generally, the Phase II results indicate that the amount of time devoted to instruction is positively associated with student learning.

A related finding is that the content covered during instruction is positively related to learning in mathematics. Work Diary-2, a performance variable involving the variety of skills covered during instruction, and Work Diary-3, a performance variable reflecting the quality of the teaching methodology for content covered, were each associated with higher outcomes in mathematics at grade 2. RAMOS-2, a performance variable reflecting the time spent by students in a variety of skill areas, was positively associated with mathematics outcomes at grade 5, but negatively associated with reading outcomes at grade 2. Hence, the positive relationship between content covered and student learning in mathematics seems to be reasonably consistent. The hint of a negative relationship in reading is difficult to explain and should be investigated further before any conclusions are formulated. It is possible that the single negative association for reading is related to the difficulties of isolating different skills taught

in reading, where instructional objectives may not be as distinct and separable as in mathematics. It may be more difficult to reliably categorize the individual skills covered during reading instruction than it is during mathematics instruction.

The Phase II results indicate that although it is advantageous to maximize the range of content covered (at least in mathematics), the instructional process should be goal-directed and focused. That is, whereas the teacher may cover a variety of objectives, each objective is a goal that requires sustained and focused instruction. APPLE-10, a performance variable that includes sustained, goal-directed teacher behaviors, was associated with higher outcomes in reading and mathematics at grade 5. APPLE-9, a performance variable involving brief, unfocused teacher behaviors, was associated with lower outcomes in mathematics at grade 2. Hence, instruction requires both breadth (content coverage) and intensity (goal-directedness).

The findings of this study also indicate that instructional interaction by the teacher is most effective when it provides the student with some relevant information. We have already seen that nonresponsiveness (APPLE-14) is not associated with either positive or negative effects. Surprisingly, instructional responses to individual students (APPLE-12, a performance variable) were associated with lower outcomes at grade 2 for reading and mathematics. This may be due in part to an associaton between frequency of individual instructional responses and the tendency of a teacher to provide unfocused instruction without goal-directedness. In any case, an examination of the results for the more specific APPLE variables indicates that the effects of instructional responses depend on the characteristics of those responses. In particular, instructional responses seem to be more effective when they have some informational component.

This conclusion is supported by several specific APPLE findings. Teacher feedback to students was associated with higher outcomes generally at both grades and subject matter areas. Working with the students and checking their work was associated with higher reading outcomes. Providing help and positive feedback (including praise) was associated with higher mathematics outcomes at grade 5. On the other hand, asking questions was associated with lower outcomes in both reading and mathematics. Discussions (where the students are expected to provide information as much as the teacher) were associated with lower reading outcomes. Teacher recognition of a student behavior, without an informational component, (e.g. acknowledging a student statement) was associated with lower mathematics outcomes at both grades.

It should be noted, however, that the Phase II results are not entirely consistent with the conclusion that informative instructional interaction is beneficial to student learning. The findings of the evaluation of videotape recordings are ambiguous in relation to this issue. For example, teacher feedback was associated with higher outcomes at grade 2 and lower outcomes at grade 5, while asking general questions was associated with higher reading outcomes at grade 2 and lower mathematics outcomes at grade 5. However, most of the teachers in Phase II were videotaped during only one day. Observational data were collected on at least two different days for each teacher and Work Diary reports were obtained for two different weeks. Therefore, it is possible that the ambiguity of the videotape results pertaining to the present discussion may be a function of the relative instability of instructional behaviors of teachers and the relative infrequency of the videotaping during Phase II. That is, behaviors such as providing feedback and asking questions may vary somewhat from day to day, so that data averaged over at least two days would be much more

generalizable than data obtained on only a single day. In any case, the majority of the Phase II data seem to indicate that instructional interaction wherein the teacher provides information is beneficial to student learning.

Student Behaviors

The Phase II findings for student behaviors are primarily concerned with the extent to which students are attentive and productive during reading and mathematics instruction. The results are somewhat difficult to interpret. Although inattentive student behavior was consistently associated with negative outcomes, attentive and productive student behavior showed no consistent pattern, being related to higher outcomes no more often than to lower outcomes.

The data on student behavior were collected primarily with the APPLE instrument, although some relevant data were also collected with RAMOS. APPLE-2, a performance variable assessing off-task (inattentive) behaviors, was related to lower outcomes for reading at grade 5 and for mathematics at grades 2 and 5. APPLE-3, a performance variable assessing on-task (attentive) behaviors, was also associated with lower outcomes, although this occurred only for mathematics at grade 5 with no effects in the other three quadrants. APPLE-1, a performance variable measuring productive behavior (good reading or mathematics performance as judged by observers), was associated with lower reading and mathematics outcomes at grade 5 and higher reading outcomes at grade 2. This confusion is furthered by results for RAMOS-2, a performance variable reflecting the amount of time spent by students practicing facts and skills. RAMOS-2 was related to lower reading outcomes at grade 2 and higher mathematics outcomes at grade 5.

The more specific variables obtained from the APPLE instrument yield somewhat more consistent results. Inattentive behaviors were consistently related to negative outcomes in mathematics, although fewer effects for inattention were found in reading. Attentive behaviors were generally associated with positive outcomes in both reading and mathematics.

Some of the descriptive data on student behaviors suggest possible explanations for the anomalies of some of these results. First of all, roughly half of the instruction in each of the four quadrants was conducted in the seatwork setting. Students also worked independently of the teacher about half of the time. In addition, approximately two-thirds of the student behaviors evoked no teacher response at all. Hence, a large proportion of student behavior, including academic work, is generally carried out without direct interaction with the teacher. The second descriptive finding is that little inappropriate behavior occurred during academic instruction, yet students were often rated as working unproductively, especially when working independently in an individual context. Hence, students may often behave appropriately, possibly with apparent attention to an assigned task, without being particularly productive.

The implication of these descriptive findings is that student attention to a task does not imply productivity and ongoing learning. This would be especially true during unsupervised seatwork. A student might sit quietly, interspersing his work with daydreaming or even unobtrusive chatter with a neighbor. This behavior might appear to be reasonably attentive. A similar problem exists for the measurement of productivity, although the descriptive data are not directly relevant here. Students could be working busily, yet without any comprehension of the task. This might continue for extended periods of time if the student is working independently. In addition, a student might work diligently on an excessively simple

task from which he has nothing to learn.

These considerations might explain the generally consistent negative relationship of inattention with learning and the inconsistent relationships of both attention and productivity with learning. That is, if a student is not attending to a task, then it is very unlikely that he is learning anything. However, the fact that a student is working attentively may not mean that he is learning, because the attention may be intermittent or at a low level of concentration. In addition, the student may be attending diligently to a task that he does not understand (as when a student systematically responds incorrectly to all the problems on a worksheet) or a task that is excessively simple. Ratings of productivity are especially vulnerable to excessively simple tasks. Students may display the highest rate of correct response production when they are working on simple tasks that they have mastered previously. Hence, high levels of attention or rated productivity may not always be associated with high student learning. The inconsistent results of Phase II data on student behavior support this conclusion.

Future research should attempt to collect extensive information on all aspects of student attention and productivity. Observation per student should be frequent enough that intermittent attention can be differentiated from constant attention. In addition, the level of difficulty of the task for the student should be determined. Rather than rate productivity at the end of the day, as did Phase II observers, data on productivity should be obtained frequently during observation in the classroom. These and other modifications of procedures for observing student behavior might yield readily interpretable relationships with learning outcomes. This, in turn, would enhance analyses of the direct effects of teacher behavior on student

behavior, in addition to the usual analyses of the effects of teacher behavior on student learning outcomes.

Despite the complexities of the Phase II findings for student behavior, there was an attempt to analyze the relationships between teacher behavior and student behavior. There was a general tendency for direct teacher interaction with students to be associated with greater student attention, while independent seatwork settings were associated with greater inattention. This suggests that the presence of the teacher results in greater levels of student attention.

It would be useful to identify the particular teacher behaviors that optimize student attention and productivity. However, this sort of relationship is probably too complex for current analyses. For example, APPLE-13, a performance variable measuring social control behaviors that are task orienting, is positively related to grade 5 mathematics outcomes. One would assume that this occurs because social control behaviors of the teacher result in increased on-task behavior for students. However, the structural analysis of the data did not show any effects of APPLE-13 (social control) for APPLE-3 outcomes (on-task behavior). Furthermore, student on-task behaviors as measured with APPLE-3 were negatively associated with grade 5 mathematics outcomes (as discussed previously). Hence, it is probably not fruitful to attempt a detailed analysis of the relationship between teacher behaviors and student behaviors, given the complexities of the learning outcomes associated with student behaviors in the Phase II analyses. The rather simple conclusion suggested above, that teacher presence is associated with greater student attention, is probably sufficient here. Future research might profitably investigate relationships between teacher and student behavior if data collected on student behavior show readily interpretable relationships with learning outcomes.

Teacher Characteristics

Phase II analyses of teacher characteristics include data on teacher expectations for students, the organizational context (district and school) within which teachers provide instruction, teacher knowledge of subject matter, teacher attitudes, and teacher aptitudes. With the exception of teacher expectations, these data were analyzed primarily in terms of their relationships with teacher behaviors, rather than in terms of their direct effects upon student learning.

Teacher expectations were analyzed in terms of their direct effects on student learning. Although expectations did prove to be related to student learning outcomes (higher expectations being associated with greater learning), the percentage of variance accounted for by expectations was relatively small (about 3 percent). Furthermore, initial student level of achievement was the best predictor of teacher expectations. Therefore, it is difficult to detemine whether the relationship between teacher expectations and student learning is due to the effects of the expectations or the effects of initial student characteristics that influence the expectations. The small relationship between expectations and final student achievement found in the Phase II data was determined after the effects of initial achievement had been removed from final achievement. Nevertheless, the strong relationship between teacher expectations and student initial achievement suggests that there may be other initial student characteristics, including academic characteristics not fully measured by the Phase II achievement tests, that influence teacher expectations and have effects on final student achievement. Therefore, it is difficult to interpret the relatively small residual association between teacher expectations and final achievement. Hence, teacher expectations should receive some examination in future research, but

Teacher Characteristics

Phase II analyses of teacher characteristics include data on teacher expectations for students, the organizational context (district and school) within which teachers provide instruction, teacher knowledge of subject matter, teacher attitudes, and teacher aptitudes. With the exception of teacher expectations, these data were analyzed primarily in terms of their relationships with teacher behaviors, rather than in terms of their direct effects upon student learning.

Teacher expectations were analyzed in terms of their direct effects on student learning. Although expectations did prove to be related to student learning outcomes (higher expectations being associated with greater learning), the percentage of variance accounted for by expectations was relatively small (about 3 percent). Furthermore, initial student level of achievement was the best predictor of teacher expectations. Therefore, it is difficult to detemine whether the relationship between teacher expectations and student learning is due to the effects of the expectations or the effects of initial student characteristics that influence the expectations. The small relationship between expectations and final student achievement found in the Phase II data was determined after the effects of initial achievement had been removed from final achievement. Nevertheless, the strong relationship between teacher expectations and student initial achievement suggests that there may be other initial student characteristics, including academic characteristics not fully measured by the Phase II achievement tests, that influence teacher expectations and have effects on final student achievement. Therefore, it is difficult to interpret the relatively small residual association between teacher expectations and final achievement. Hence, teacher expectations should receive some examination in future research, but

should not be considered a strong candidate to account for student learning.

Teacher characteristic variables other than expectations were analyzed primarily in terms of their relationships with teacher behaviors (using path analysis procedures), rather than in terms of their direct relationship with student learning. There is a large body of well-publicized research relating teacher expectations directly to student outcome measures. Therefore, it is appropriate that this study is consistent with that previous research. However, it is also true that a logical analysis of the relationship between teacher characteristics and student learning would generally conceive of this as an indirect relationship, wherein the characteristics influence teacher behaviors in the classroom, which in turn influence student learning. The structural model used for the path analyses of Phase II takes this view, analyzing the relationships between teacher characteristics and teacher behaviors. Teacher behaviors may, in turn, be related to student learning. However, it is difficult to interpret the complex paths from teacher characteristics through teacher behaviors to student learning, especially with the additional intermediate variable, student behavior, between teacher behaviors and student learning. Previous discussion has illustrated the complexities of interpreting paths from teacher behaviors thorugh student behaviors to student learning. Clearly, it is even more difficult to interpret paths starting with teacher characteristics and leading, eventually, to student learning.

In light of the complexities mentioned above, the present discussion will focus on the relationship between teacher characteristics and teacher behaviors, without attempting to follow an associative path to student learning. Therefore, no conclusions will be drawn regarding the instructional effects of teacher characteristics (other than expectations) in terms of student learning. Rather, this discussion will consider which teacher

characteristics are most promising for future research in terms of their relationships with teacher behaviors as indicated by Phase II findings.

The organizational context of instruction, involving district and school characteristics, showed some relationships to teacher behavior,but the number and strength of these relationships were insufficient to warrant their consideration as promising variables for future research. Variables related to organizational climate were generally weak predictors of teacher behavior. Organizational structures, such as team teaching and the use of aides, were also generally unrelated to teacher behaviors. Even where relationships were found, the Phase II report points out that it is difficult to determine whether the organizational context variable had a direct effect on the associated teacher behavior. An alternative hypothesis would be that the teachers who typically show that behavior were selected by administrators for employment in the districts (and/or schools) characterized by that organizational context. For example, it was found that teachers within an organizational context characterized by centralized decision-making engaged in more whole-class instruction with sustained interactions. It may be that district administrators who prefer central control of decision-making tend to encourage teachers to maximize their direct control over the entire class by using whole-class instruction with sustained interactions. However, it is also possible that the same district administrators would simply select for employment teachers who show a preference for that style of instruction. Therefore, even where organizational variables are related to teacher behaviors, clear interpretation is difficult. As a result, organizational variables should not be considered strong candidates for future research.

Teacher knowledge of subject matter and methodology variables were not strongly related to teacher behavior variables. Few of the path coefficients

in this domain showed effects for teacher behavior variables. A related finding from separate analyses was reported in the special study on the use of a videotape test of teacher diagnostic skills in reading. Surprisingly, positive scores on this test were related to negative student reading outcomes. These results indicate, as the Phase II report recommends, that future research should modify or replace the measures of teacher knowledge (and diagnostic skills) used here. It seems unlikely that knowledge of the academic content being taught and the instructional methods related to that content should be unrelated to the effectiveness of instruction. However, additional instrument development will be necessary if such relationships are to be identified.

Teacher attitude variables showed some relationships with teacher behaviors, but not with the frequency and magnitude that would warrant emphasis in future research. However, one interesting relationship that was identified was the association of higher teacher aspirations with the use of more whole-class instruction and less independent seatwork. It is possible that teachers with higher aspirations feel a need for the greater control over student behavior that is generally possible with whole-class instruction. An additional finding is that aspirations tend to be positively associated with the same teacher behaviors with which aptitude variables are positively associated. The Phase II report suggests that future research might investigate teaching styles characterized by combinations of attitude (such as teacher aspirations) and aptitude variables (discussed below).

Teacher aptitude variables were strongly related to teacher behaviors as indicated by the frequency and magnitude of the path coefficients relating these aptitudes and behaviors. Several interesting relationships were found. At grade 2, more flexible teachers tended to use a great deal

of whole-class instruction and little independent-student instruction. More flexible teachers also tended to do less organizing and supervising (APPLE-8). Field independence showed a negative relationship with the use of social control for task orientation (APPLE-13) at both grades and subject matters. Another interesting finding was the positive relationship between verbal fluency and the quality of teaching methodology (Work Diary-3) in reading, contrasted with the negative relationship between reasoning ability and the quality of teaching methodology in mathematics.

Aptitude variables are complex in any analysis. The variety of relationships represented in these analyses only adds to that complexity. Whereas tentative explanations could be offered for the examples cited above, such explanations would not be simple or obvious. Furthermore, it is difficult to identify policy applications of these findings. These aptitude and behavior relationships are not readily amenable to teacher training procedures. In addition, one would hesitate to select teachers according to measured aptitude without both replication of these relationships and a much clearer understanding of their consequences for student learning.

Nevertheless, future research should attempt to further develop the Phase II finding that teacher aptitudes are related to teacher behaviors. The previously mentioned strategy of describing teaching styles in terms of combinations of attitude and aptitude variables would be one way of building on the present findings. Another possibility is that teachers could be asked to give their rationale for the teaching style that they use. These rationales might then be interpreted in terms of the attitude and aptitude variables discussed here. Structured interviews might elicit these rationales with consideration of aspirations, flexibility, field independence, and other attitude or aptitude variables. This approach might clarify the relationship between these teacher characteristics and instructional behaviors.

Student Characteristics

Student characteristics, including SES, aptitude, expectations, and attitude, were examined for their possible value in accounting for student learning. As in the other Phase II analyses, student learning is conceived as that part of a spring test score that is not accounted for by a fall test score. (This is generally calculated in terms of gain scores or residual spring scores). However, it should be noted that fall test scores account for 70 to 80 percent of the variance in spring scores for Phase II data. Hence, the most important student characteristic associated with the level of achievement reached by the end of the year is the student's level of achievement at the beginning of the year. Relative to initial level of achievement, other student characteristics are not strong predictors of achievement level at the end of the year.

Student SES (socioeconomic status) was actually just one of many student background characteristics examined in Phase II. Other student background variables included educational experience (such as preschool), non-English linguistic experience, sex, and race. SES proved to be much more highly correlated with both fall and spring total reading and mathematics scores than any of the other student background variables. This was true with several different procedures for rating SES, including composites from teacher and parent questionnaires and simple quantifications of parental education. However, neither SES nor any of the other background variables was strongly related to student learning. It may be concluded, therefore, that SES is the most important student background variable, but that even SES is not important in explaining differences in student learning over the course of a school year. Future research should probably consider SES

simply because it is the most powerful student background variable. However, the Phase II results suggest that SES should be used as a descriptive variable, rather than as an analytic variable for explaining student achievement.

Student aptitude did show some relationships to student learning, but only in a relatively low proporation of the analyses conducted. Furthermore, the association between aptitude and learning was not always readily interpretable. For example, verbal aptitude was negatively related to reading outcomes at grade 2 and unrelated to outcomes in the other three quadrants. Field-independence showed a strong positive relationship with mathematics outcomes at grade 2, but no relationship with mathematics at grade 5 and small positive relationships with reading outcomes at grades 2 and 5. It is reasonable to find a relationship between a non-verbal aptitude variable, such as field-independence, and mathematics outcomes. However, it is difficult to understand why this relationship would hold only at grade 2, showing a stronger relationship to reading outcomes at grade 5 than to mathematics outcomes. These results, showing aptitude to have a low and somewhat complex relationship with student learning, suggest that student aptitude variables should not receive a great deal of emphasis in future research on teaching.

Student expectations (for student achievement) showed a moderate relationship with learning outcomes. However, this relationship is difficult to interpret. Relationships between fall expectations and learning occurred only in reading and were always negative. Spring expectations were related to both reading and mathematics outcomes, but this relationship was weak and was equally divided between positive and negative coefficients. Therefore, student expectations show little promise for future research on teaching.

Student attitudes toward reading and mathematics also showed only moderate relationships with learning outcomes. However, attitudes showed the interesting pattern of negative relationships between fall attitudes and learning contrasted with positive relationships for attitudes measures in spring. The majority of the fall coefficients were negative and all of the spring coefficients were positive. This pattern suggests that change in attitude over the year might be related to learning. In addition, there was some evidence in the Phase II findings that some teacher behaviors related to higher learning outcomes may be related to negative changes in attitude. Therefore, it is probably important that future research examine change in attitude as a check to be sure that the teaching behaviors associated with higher learning outcomes are not also associated with negative attitude effects.

Possible Reanalyses of Phase II Data

The Phase II report presents and discusses a wide variety of alternative analyses that were conducted in the Phase II data. These include simple correlations, multiple and stepwise regressions, and path analysis using a structural model. In addition, analyses were conducted for both mean gain and residual scores, as well as for both the total group of classes and a discriminant group of classes (those with ten or more students with complete data). There was also an innovative analysis of teacher effects according to the slope and intercept of the within class regression of spring scores on fall scores. The scope of these analyses is exhaustive. Additional analyses of the same data in its current form are not advisable, in that the analyses reported explore all aspects of these data as they are presently conceived. However, secondary analysis might reconceptualize the data so as to examine relationships other than those that were within the domain

of the Phase II report. Two such possibilities are suggested below.

The tremendous range and variety of data collected during Phase II, and the time constraints within which the primary analyses were conducted, required the development of composite variables on the basis of statistical and conceptual relationships between specific variables. Secondary analysis might examine different composite variables than those currently analyzed in those cases where an alternative conceptual framework appears reasonable. For example, a large number of categories of behaviors, events, and context features were recorded during observation with the RAMOS instrument. These data were combined to create four variables, one for each of the following categories: 1) the teacher's time in a variety of instructional roles, 2) student time in a variety of instructional aims (primarily practicing facts and skills), 3) instructional time with a variety of materials, and 4) instructional time in a variety of activities. Whereas these are important domains, it would also be interesting to examine the RAMOS data using variables within the domains of teacher time, student time, materials, and activities. In some cases it might be advisable to combine subcategories from within different domains. For example, within the domain of teacher time one might compare time in direct instruction with time facilitating or monitoring instruction. One might also combine across categories, perhaps examining time in direct instruction during lecture or discussion and time facilitating or monitoring during seatwork.

Redefinitions of composite variables might also be considered for Work Diary data. For example Work Diary-1 might be divided into two variables: 1) time spent preparing for teaching and 2) time spent teaching reading and mathematics.

A second possibility for secondary analysis is to attempt alternative organizations of the data for the purpose of examining other relationships

between variables. For example it might be possible to analyze the observational data collected on target students (eight per class) with both the APPLE and the RAMOS observation instruments. This would allow for consideration of the different instructional events and conditions received by different students within the same class. Both between subjects (across all classes) and pooled within-class analyses might be conducted. The pooled within-class analyses would eliminate between-class effects. This approach might be especially useful for examining the effects of student behavior, such as productivity, inattention, and attention.

It should be noted that the possible secondary analyses discussed above would require major commitments of resources. Far West Laboratory is not prepared to conduct analyses of this magnitude for the Phase II data. Less comprehensive, supplementary analysis is not required or advisable given the range and variety of analyses included in the Phase II report (as discussed above).

Methodological Implications

The Phase II report yields several valuable methodological implications for conducting research on teaching. This is particularly important, because research in any field is an ongoing process where every answer attained reveals several new questions. Therefore, methodological advances can be applied to an increasing number of questions that must inevitably be investigated.

Perhaps the most important implication is that future research should obtain and examine relatively simple data sets with a minimum number of variables, each of which should be selected for the maximum impact upon the object of investigation. The very large number of variables included in the Phase II data set was appropriate for an initial, exploratory study.

Future research, however, should use fewer variables so that statistical complications, such as over-determination and multicollinearity can be reduced or eliminated. In addition, the use of fewer variables should simplify and enhance data interpretation.

Simplification of the data set required for a particular investigation must involve an organizing strategy so that critical domains of variables and relationships are not excluded. One possible strategy, recommended by the Phase II report, is the examination of patterns of teaching behavior, rather than single, isolated behaviors. Patterns might be conceptualized in terms of their functional relationships for instructional purposes. This approach would have several consequences for future research on teaching. For example, teacher preparation for instruction and interacting teacher behaviors during instruction would ba analyzed as separate variables. Preparation performs a planning or decision-making function, whereas interaction during instruction involves the function of implementing previous decisions. In addition, the degree of teacher-involvement in instruction (from teacher-led to independent student) would be distinguished from the degree of individualization of instruction. Teacher-involvement refers to the degree to which the teacher performs a direct, interactive role within a particular instructional setting. Individualization is determined by the functional referent of the objectives and tasks of instruction. That is, objectives and tasks may be designed to perform a function for an individual student, a group of students, or an entire class.

The conceptualization of variables in terms of functionally-related patterns can also be applied to student behaviors. Rather than assessing student attention or productivity, one would assess attention or productivity in terms of a particular task, described in terms of the test or subtest (content area) for which that task is relevant. That is, the student

behavior would be characterized in terms of the academic function that it performs. Attention to a page of addition problems does not perform the same function as attention to word problems on place value. One might also record the difficulty level of the task for the particular student involved in that task. This might differentiate between tasks that function to provide review (low difficulty), introduce new material (moderate difficulty), or simply frustrate the student (excessively high difficulty).

Another implication of the Phase II findings is that data analysis procedures should be relatively simple and straightforward in future research. Studies of teaching are still at an early stage of development, where sophisticated statistical procedures are probably not useful. For example, the effective use of path analysis requires the ability to identify meaningful relationships at various points throughout a complex structural model. This resulted in several complications of the interpretation of the Phase II path analysis. One such instance was the finding that student attentiveness was negatively related to learning outcomes for grade 5 reading. This unexpected result makes it very difficult to interpret any teacher behaviors that are related to grade 5 student attentiveness, using path analysis. Therefore, the less sophisticated regression techniques are probably more appropriate here. These techniques allow the researcher to isolate anomalous findings rather than attempting to integrate them into a complex analysis of an entire data set.

Phase II also has methodological implications for the collection of data on classroom instruction. It is suggested in the Phase II report that future observation with time-based instruments, such as RAMOS, should cover as many days as possible, evenly distributed throughout the instructional period under study. In addition, it is recommended that observation should cover as much of the school day as possible. These procedures would

greatly enhance the generalizability of the data collected. In addition, it would be possible to estimate the total time for an instructional term that should be assigned to a given observation category. This avoids the necessity of analyzing proportions of observed time that are assigned to a given category, wherein differences in the total amount of instructional time cannot be considered.

Another methodological implication for observation is that the number of categories for which observation data is collected should be reduced to a minimum. The requirements of analysis procedures, not to mention comprehensible interpretation of the results, necessitate the collapsing of many categories to a manageable few. This was the case in Phase II. If most of the collapsing process were carried out in advance, then observers would have far fewer categories to process. As a result, the reliability of observation would probably be enhanced without sacrificing any of the distinctions required for analysis purposes.

Recapitulation and Conclusion

Phase II findings have been discussed in terms of four categories of variables: 1) teacher behaviors, 2) student behaviors, 3) teacher characteristics, and 4) student characteristics. Consideration was also given to possible reanalyses of the data and methodological implications of the study. As a means of summarizing the findings discussed here, the major conclusions are listed below.

The largest number of variables suggested by the Phase II findings is in the category of teacher behaviors. The finding that quantity of teacher preparation for instruction is unrelated (probably even negatively related) to student learning suggests that the quality of a teacher's preparation may be the more important variable. It was found that the teacher's direct

involvement in instruction is related to greater student learning. There was some indication, though not conclusively, that the individualization of instruction (assuming a constant level of direct teacher involvement) is related to higher student learning outcomes. The use of greater variety of materials was usually, though not always, related to lower levels of student learning. This may be due to a detraction from the structure and focus of instruction associated with a greater variety of materials. Management behaviors were associated with both positive and negative outcomes. It may be important to distinguish between management behaviors that are the result of the teacher's general task orientation and those that are the teacher's response to an immediate behavioral disorder in the classroom. The quantity of time devoted to reading and mathematics instruction was positively related to student learning in those areas. In addition, the range of content areas covered in mathematics was positively related to mathematics achievement growth. Higher levels of focus and goal-directedness manifested in instruction were associated with greater student learning. Instructional responses of the teacher to students where there was a direct informational component (as opposed to asking questions or leading a discussion) were associated with higher student learning outcomes.

Although student behaviors showing inattention were related to lower learning outcomes, there was no consistent relationship between student attention or productivity and student learning. It was concluded that other features of student task involvement, such as the difficulty of the task for the student, should be investigated in conjunction with attention. It was also suggested that data on student attention should be based upon relatively large amounts of observation. The complexity of many of the relationships between student behavior and student learning make it difficult

to interpret relationships between teacher behavior and student behavior. However, it can be said that the presence of the teacher was associated with increased student attention to the task.

The most promising domain of teacher characteristics was teacher aptitude. Flexibility and field independence seemed to be the most important aptitude variables. Teacher attitudes, especially aspirations, did show a moderate relationship with teacher behavior. It was suggested that aptitudes and attitudes could be studied in relationship to clusters of teaching behaviors (styles) by investigating the teacher's rationale for his/her teaching style. Structured interviews might elicit these rationales with consideration of flexibility, field independence, and aspirations.

Student SES (socioeconomic status) was related to achievement, but it did not appear to be as useful for explaining learning (where initial achievement is "removed" from final achievement). It was concluded that SES should be examined for descriptive, rather than analytic, purposes. Student aptitude, expectations and attitude each showed only a few associations with learning. However, whereas aptitude and expectations showed contradictory or inconsistent relationships, student attitude displayed an interesting pattern. Attitude measured in the fall was negatively related to learning over the year, while attitude measured in the spring was positively related. It was concluded that change in attitude over the year should be investigated.

The comprehensive scope of the analyses conducted and reported for the Phase II data obviate any consideration of supplementary reanalyses. However, secondary analyses might be considered, reconceptualizing the data so as to examine relationships other than those within the domain of the Phase II report. One such possibility would be the redefinition of

some of the composite variables so that "new" variables could be examined (using the same data). Another possibility would be to reorganize the data for analysis so that observational data on target students could be examined, using the individual student as the unit of analysis.

Several important methodological implications can be drawn from the Phase II findings. It was suggested that the variables investigated and the data analysis procedures employed should be simplified so that complications of analysis and interpretation can be avoided. It was suggested that variable domains should be conceptualized in terms of patterns of behavior that have functional relationships within the instructional process. It was also suggested that observation should cover as much of the school day as possible, sampling days at frequent intervals throughout the term of the study. In addition, observation should be conducted with the minimum number of categories required for analysis purposes, thereby enhancing the reliability of the observational data.

Despite the large number of variables and conclusions outlined above, it is clear that these do not exhaust the possible interpretations and conclusions that could be drawn from the Phase II findings. The scope, complexity, and richness of the Phase II data preclude any pretense that the present review might be comprehensive. Nevertheless, it is clear that those variables and conclusions discussed here will be extremely valuable in future research. It is the responsibility, and the opportunity, of future research to build upon these findings. Phase III-B of BTES will be a first step in this direction, incorporating both the general conceptual implications and the specific findings for particular variables into the design and instrumentation of that research.[3]

[3]See the Proposal for Phase III-B of the Beginning Teacher Evaluation Study and Technical Note I-2 (in press) for further elaboration of the relationship between Phase II findings and the Phase III-B design and instrumentation.